ELSEVIER

An Application of a Weighting Method to Adjust for Nonresponse in Standardized Incidence Ratio Analysis of Cohort Studies

R. SOWMYA RAO, PHD, ALICE J. SIGURDSON, PHD, MICHELE MORIN DOODY, MS, AND BARRY I. GRAUBARD, PHD

PURPOSE: Cohort studies often conduct periodic follow-up interviews (or waves) to determine disease incidence since the previous follow-up and to update measures of exposure and confounders. The common practice of excluding nonrespondents from standardized incidence ratio (SIR) analyses of these cohorts can bias the estimates of interest if nonrespondents and respondents differ on important characteristics related to outcomes of interest. We propose an analytic approach to reduce the impact of nonresponse in the analyses of SIRs.

METHODS: Logistic regression models controlling baseline information are used to estimate the propensity, or the probability of response; the reciprocals of these propensities are used as weights in the analysis of risk. This is illustrated in the analysis of 15 years of follow-up of a cohort of US radiologic technologists after an initial interview to assess the risk at several cancer sites from occupational radiation exposure. We use information from the baseline survey and certification records to compute the propensity of responding to the second survey. SIRs are computed using Surveillance, Epidemiology, and End Results (SEER) cancer incidence rates. Variances of the SIRs are estimated by a jackknife method that accounts for additional variability resulting from estimation of the weights.

RESULTS: We find that, in this application, weighting alters point estimates and confidence limits only to a small degree, thus providing reassurance that the results are robust to nonresponse. This indicates that results from the analyses excluding the missing data may be slightly biased and weighting helps in reducing the nonresponse bias.

CONCLUSION: This method is flexible, practical, easy to use with existing software, and is applicable to missing data from cohorts with baseline information on all subjects.

Ann Epidemiol 2005;15:129–136.

KEY WORDS: Nonresponse, Cohort, Weighting.

INTRODUCTION

Response rates have been declining over the years in population surveys and cohort studies (1, 2). Despite several studies addressing issues of recruiting, tracking, and retaining cohort members (3–5), a 100% response rate is typically unachievable. Failure to account for nonresponse can lead to incorrect inference, the extent of which can depend on the amount of missing information, the nature of the missing data mechanism, and the estimates of interest (6).

In this article we address the problems due to nonresponse when conducting standardized incidence ratio (SIR) analyses of multiple-wave cohort studies, where individuals are contacted at each follow-up (wave) to determine disease incidence since the last wave. Subjects in the cohort may not respond to one or more of the follow-ups, resulting in the loss of important disease information. This is in contrast to cohort studies without direct follow-up contacts where disease incidence is obtained from a disease registry system and, therefore, disease status is known for essentially the entire cohort regardless of response at each wave. However, even these types of cohort studies can have some missing disease outcome data when subjects move out of the areas covered by the registry system.

In epidemiological cohort study analyses that prospectively collect time-to-disease data, traditional survival analyses treat missing data due to study dropout or nonresponse to follow-up as censored observations. For example, in a two-wave cohort study (with a baseline and a follow-up interview) subjects with nonresponse to disease incidence would be excluded from a SIR analysis. This would be equivalent to a "complete-case analysis" and would explicitly assume that the missing data is missing completely at random (MCAR) (6). This is often done despite there being available covariates measured at baseline (e.g., age, prior medical conditions) that are predictive of response to the survey and disease incidence. It is well known that

From the Biostatistics Branch (R.S.R., B.I.G.), and Radiation Epidemiology Branch (A.J.S., M.M.D.), Division of Cancer Epidemiology and Genetics, National Cancer Institute, DHHS, Bethesda, MD.

Address correspondence to: Dr. R. Sowmya Rao, Biostatistics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, 6120 Executive Boulevard, EPS 8047, MSC 7244, Bethesda, MD 20892-7244. Tel.: (301) 435-3998; Fax: (301) 402-0081. E-mail: raos@mail.nih.gov

Received June 23, 2003; accepted May 20, 2004.

360 Park Avenue South, New York, NY 10010

1047-2797/05/$–see front matter
doi:10.1016/j.annepidem.2004.05.007

AEP Vol. 15, No. 2
February 2005: 129–136

Selected Abbreviations and Acronyms

ARRT = American Registry of Radiologic Technologists
SIR = standardized incidence ratio
SEER = Surveillance, Epidemiology, and End Results
MAR = missing at random
MCAR = missing completely at random
ROC = receiver operating characteristic

complete-case analysis can be potentially biased (6). In this article, we deal with missing outcome information and how the results can be biased if the response depends on observed (or unobserved) covariates.

Weighting and imputation are two methods available to adjust for nonresponse by using important baseline covariates to conduct missing at random (MAR) analyses (6–9). Weighting adjusts the contribution of respondents to represent comparable nonrespondents, while imputation assigns values to nonrespondents based on the respondents. In longitudinal studies, covariates measured at baseline for the entire sample are used by weighting methods to model the probability of missing in the follow-up (7, 8) or used by imputation methods to model the assignment of the missing values. The validity of both types of methods depends on the missing at random assumptions conditional on the baseline covariates (6).

When there is a reasonable amount (> 10%) of missing information, one is not sure how the estimates will be affected. One way to evaluate the effect is by conducting an MAR analysis which adjusts for the nonresponse and compare the results with those obtained from a "complete-case" analysis.

In this article we propose a weighting approach used in survey research to reduce bias from nonresponse when subjects are repeatedly contacted in panel surveys (10, 11). We model the probability (propensity) of being a respondent as a logistic function of predictor variables and compute weights as the inverse of the predicted propensities (2, 12–17). These weights are used to adjust for nonresponse. We demonstrate the usefulness of this weighting method by applying it to a two-wave cohort study in which an SIR analysis was conducted (18). We suggest a method for estimating standard errors for the weighted estimates and for constructing confidence intervals. This approach may be extendable to analyses of absolute risks but this issue is beyond the scope of this work.

METHODS

Cohort Description

In 1982, a cohort of 143,517 radiologic technologists was assembled from the computerized certification files of the American Registry of Radiologic Technologists (ARRT) (19–21). To be eligible for cohort entry, the technologist must have been certified by the ARRT for 2 years or longer between 1926 and 1982, and must have resided in the US or its territories. A 12-page questionnaire (Survey 1) was mailed to over 132,000 radiologic technologists who were presumed to be alive between 1983 and 1989 and had valid addresses. The questionnaire included items about work history as a radiologic technologist, radiation protection methods, lifestyle characteristics, demographic factors, and health outcomes, including cancer. The response rate to the baseline survey was 68% (n = 90,305). Of these 90,305 subjects, 3103 subsequently died, and 70,859 responded to a second questionnaire (Survey 2) administered between 1995 and 1998. Information was again elicited on various health outcomes, including cancer site and date of diagnosis. This study was approved by the human subjects review boards of the National Cancer Institute and the University of Minnesota.

The population analyzed in this article is not identical to the population described elsewhere (22) because we imposed slightly different start dates for inclusion. Thus, for example, we have excluded 3 subjects who died before our start date (January 1, 1983), resulting in 90,302 subjects rather than 90,305 (22), who we consider to have responded to Survey 1, and present the weighted results for the same. Any report of cancer was counted as an observed case. To supplement incidence histories from Survey 2, we also ascertained deaths due to cancer in the interval from death certificates or using National Death Index Plus system.

This cohort has two kinds of nonrespondents: 1) subjects who did not respond to Survey 1, and 2) subjects who responded to Survey 1 but not to Survey 2. Excluding the subjects in category 1 might affect the generalizability of results to the entire cohort but will not affect the internal validity of the results. In this analysis we only adjust for the nonrespondents in category 2.

Statistical Analysis

Basic Approach and Simple Example. We first illustrate our approach using a simple hypothetical example. Consider a cohort of 1 million men and 1 million women who responded to the baseline questionnaire, and are being followed to assess the risk of a certain type of cancer, C. Suppose, 30% of the men and 70% of the women respond to the follow-up survey mailed after 1 year. Let us assume that 15,000 of the respondents (10,000 men and 5000 women) report having C and that the expected number of type C cancers from an external source is 25,000 (5000 men and 20,000 women). If we ignored the 1 million nonrespondents (700,000 men and 300,000 women), we would have the

$$SIR = \left(\frac{Observed}{Expected}\right) = \left(\frac{15{,}000}{25{,}000}\right) = 0.6$$

If we assume that for each gender the outcomes are MAR, then this estimate will be biased because the response rate depends on gender and the cancer rates also are associated with gender. We can adjust for the nonresponse by appropriately weighting the responses. We do this by first computing the propensity (probability) of response as 0.30 for men and 0.70 for women. The weight for the respondents is the reciprocal of the propensities, 3.3333 for men and 1.4286 for women. If more than one variable is associated with response (e.g., age, education, income), we can fit a logistic regression using all these variables to model and compute the propensities. The estimated number of type C cancers after weighting for nonresponse is 33,333 (= 10,000*3.333) among men and is 7143 (= 5000*1.4286) among women. The expected number of type C cancers after weighting for nonresponse is 16,665 (= 5000*3.333) among men and is 28,572 (= 20,000*1.4286) among women. Now, we re-compute the SIR as

$$SIR = \left(\frac{\sum_{men} Observed*Weight + \sum_{women} Observed*Weight}{\sum_{men} Expected*Weight + \sum_{women} Expected*Weight}\right) = \left(\frac{40{,}476}{45{,}237}\right) = 0.9$$

As we can see from this simple example of an SIR analysis when response depends on the same factors that are associated with the outcome, weighted analyses, where the weighting is by the inverse of the propensity of response missingness, can be considerably different from the unweighted analysis; by excluding the nonrespondents we bias the estimate of SIR. In this example, the weighting gives unbiased estimates of number of observed and expected cases in the population being considered. In other words, the weighted SIR will on average be equal to the "truth" where as the unweighted SIR will not. In reality, the nonrespondents might vary from the respondents on more characteristics than gender and more so by various risk factors for the disease of interest and excluding them could bias the results. Below, we illustrate our approach with a more complicated real dataset.

Weight Estimation

Information from the ARRT certification records and from Survey 1 was used to form baseline covariates for the entire cohort of 90,302 subjects. For each gender, we performed a step-down (or backward) logistic regression analysis of the baseline covariates and two-way interactions between the covariates to obtain a model to predict the probability (propensity) to respond to Survey 2. Separate logistic models were fit for each gender because the relationship between the different covariates and response varied by gender. Also, step-down selection was used because it produces more parsimonious models that reduce variability of the ultimate sample weights derived from the propensities predicted using these models; increased variability in the weights usually leads to increased variances in the analysis (17).

Of the 90,302 subjects who responded to Survey 1, a total of 3100 died prior to Survey 2, and 16,343 (18%) subjects who were still alive did not respond to Survey 2. The weight for an individual who responded to Survey 1 was calculated as the inverse of the estimated propensity of response. Because persons who died between Survey 1 and Survey 2 have their disease status essentially completely assessed from death certificates they were assigned a weight of one. The weights for the male technologists ranged from 1.00 to 4.95 (median, 1.36; coefficient of variation, 33.72) and the weights for the female technologists ranged from 1.00 to 5.96 (median, 1.23; coefficient of variation, 30.63). Since extreme weights can decrease efficiency, we trimmed the upper 5% of our weights (to 2.95 for the male technologists and 2.17 for the female technologists) (23, 24). The weights were then rescaled (i.e., total sample size/sum of the weights) to sum to the sample size.

Parameter Estimation

We computed the SIRs (unweighted) as the ratio of the observed number of cases to the expected number of cases, where the expected number are derived from the Surveillance, Epidemiology, and End Results (SEER) cancer rates (http://www.seer.cancer.gov). We computed the weighted SIR for a specific cancer of interest using the following formula:

$$SIR_w = \frac{O_w}{E_w} = \frac{\sum_j w_j C_j}{\sum_a S_a \sum_j w_j T_{aj}}$$

where w_j = weight associated with the j^{th} respondent in cohort; $C_j = 1$ if j^{th} respondent had the cancer of interest and 0 otherwise; T_{aj} = person-years for the j^{th} respondent in the age-sex-race-calendar-year-specific cell "a" and free of the cancer of interest; and S_a = SEER rate for the cancer of interest in the age-sex-race-calendar-year-specific cell "a". When the w_j are all equal to the same value then we get the unweighted SIR.

Variance Estimation

Because weighting was used to estimate the SIR_w, standard formulas for variance calculation of unweighted SIRs could

AEP Vol. 15, No. 2
February 2005: 129–136

not be used. A few of the methods that can be used to obtain approximately unbiased variance estimates are grouped jackknife, bootstrap, and Taylor linearization (referred to as the delta method) (17). Taylor linearization requires derivation of analytical formulas and specialized software. However, both the bootstrap and grouped jackknife, which are based on replicating the analysis with different subsets of the data, are easily implemented with standard software but computer intensive. Since the bootstrap method is more intensive because it requires more replications of the sample, we used the delete-one-group-jackknife method to compute the variances. The version of the delete-one-group-jackknife that we used is described as follows: We randomly divide the dataset into k (= 30) approximately equally sized groups. We use the weights that were calculated for the overall dataset to obtain the $SIR_w, \hat{\theta}$. Corresponding to each of the 30 groups, we create 30 data sets by leaving out the observations from one group at a time. Using each of these 30 datasets, we compute a different set of weights by re-estimating the logistic regression model for the propensity of response. We then compute weighted SIRs,$\hat{\theta}_{(i)}$, $i = 1, \ldots, k$, for each of the 30 datasets using their respective weights. The jackknife variance estimator is computed as

$$var(\hat{\theta}) = \frac{(k-1)}{k}\sum_{i=1}^{k}(\hat{\theta}_{(i)} - \hat{\theta})^2$$

The variance estimates obtained using 30 groups were very similar to the estimates obtained using 50 groups (data not shown). Hence, we suggest that the analyst save computing time by using only 30 groups.

We used this variance to compute Wald-type confidence intervals. When the number of observed cancers is greater than zero and less than 51, we propose a modification to the (1-α) level confidence interval, which accounts for the additional variability due to the weighting of the SIR estimator (see Appendix) (25). When the number of observed cancers was zero, we ignored the weighting, and calculated the upper confidence limit using the exact Poisson formula given by Liddell (26). We conducted limited simulations to evaluate the coverage properties of the proposed confidence intervals for small number of events with expected number of events ranging from 1.2 to 38.4. In these simulations we used weights that were correlated with the outcome, i.e., the weights were informative and changed with the probability of events. We found that the coverage of the proposed 95% confidence intervals were close to 95% for larger number of events, $>$ 4.8 events, and tended to be $>$ 95% for the smaller number of events ($\leqslant$ 2), which was reassuring (data not shown).

RESULTS

Table 1 displays the distribution of selected characteristics for the respondents and nonrespondents to the second survey as well as the subjects who died between the two surveys. Respondents were proportionately more likely to be female, Caucasian, younger, and certified as a radiologic technologist for a longer period ($>$ 20 years). A higher proportion of respondents to Survey 2 were married, never smoked, and among the women, were users of oral contraceptives (data not shown).

The logistic regressions were significantly predictive of response to Survey 2. The variables predictive of response to Survey 2 among all the male technologists were region, race, and number of years certified. Additional variables that predicted response among Survey 1 responders who are male were marital status, smoking status, and an interaction term for year first worked by marital status. The variables predictive of response to Survey 2 among all the female technologists were region, race, number of years certified, and an interaction term for race by region. Additional variables that predicted response among female Survey 1 responders were marital status, smoking status, ever use of oral contraceptives, and three interaction terms: number of years certified by marital status, number of years certified by ever oral contraceptive use, and year first worked by region. These variables and interactions were used in the logistic regression analysis to compute the response propensities.

Table 2 displays selected SIR estimates and confidence intervals obtained using complete case analysis and weighted analysis. Although there were a few moderate differences between the nonresponse adjusted analysis and the complete-case analysis, in general, the two sets of estimates did not differ appreciably. Also, the widths of the confidence intervals did not change much between the two sets of SIR estimates. This indicates that there was little loss in efficiency due to the weighting adjustment. The application of the weighting method to all 64 cancer sites showed that compared with the complete-case analysis, the weighted SIRs were reduced for 52 of the 64 cancer sites (by 1–23%) and increased for 9 sites (by 1–3%). The weighted SIRs for all cancer sites are presented elsewhere (18).

DISCUSSION

We demonstrate the response propensity weighting method to adjust for nonresponse in the computation of SIRs in a multi-wave cohort study of radiologic technologists. We also describe methods for estimating confidence intervals of SIRs when the data is weighted for nonresponse.

In contrast to our hypothetical example, in the real example we find that the weighting alters the point estimates and confidence intervals only to a small degree,

TABLE 1. Characteristics of survey respondents and nonrespondents in the US Radiologic Technologist Cohort*, 1983–1998, who were alive in 1983, and responded to Survey 1 (n = 90,302)

	Total Cohort(n = 90,302)		Responded to Survey 2 (n = 70,859)	Nonrespondents (n = 16,343)	Died before Survey 2 (n = 3100)
Characteristic	N	%	%	%	%
Sex					
Female	69,523	77.0	78.7	72.8	59.8
Male	20,779	23.0	21.3	27.2	40.2
Race					
Caucasian	85,629	94.8	95.6	92.1	91.4
Others	4673	5.2	4.4	7.9	8.6
Birth year					
Before 1925	4630	5.1	3.6	4.2	44.6
1925–1934	7551	8.4	8.3	6.1	21.2
1935–1944	18,749	20.8	21.1	19.7	17.5
1945–1954	41,169	45.6	46.3	48.8	13.5
1955 +	18,203	20.1	20.7	21.2	3.1
Region of residence†					
Northeast	21,545	23.9	23.9	24.3	20.4
Midwest	26,802	29.6	30.4	27.1	26.2
South	24,925	27.6	26.9	30.5	28.5
West	17,028	18.9	18.8	18.1	24.9
First year certified as radiologic technologist					
Before 1940	346	0.4	0.2	0.4	5.6
1940–1949	2344	2.6	2.0	1.9	19.9
1950–1959	10,776	11.9	11.6	9.2	34.0
1960–1969	25,878	28.7	29.2	27.0	24.2
1970 +	50,958	56.4	57.0	61.5	16.3
Number of years as radiologic technologist					
< 10	3110	3.4	2.8	6.2	5.3
10–19	8624	9.6	7.7	15.6	19.0
20–29	47,540	52.6	53.8	51.5	31.7
30 +	31,026	34.4	35.7	26.7	44.0

*Some percentages do not add up to 100 due to missing values.
†Based on US Census Bureau definition for geographic region: Northeast = CT, ME, MA, NH, RI, VT, NJ, NY, PA, DE, DC, MD; Midwest = IL, IN, MI, OH, WI, IA, KS, MN, MO, NE, ND, SD; South = FL, GA, NC, SC, VA, WV, AL, KY, MS, TN, AR, LA, OK, TX; West = AZ, CO, ID, MT, NV, NM, UT, WY, AK, CA, HI, OR, WA.

thus providing reassurance that the results are likely robust to nonresponse. Does this mean we can conduct a complete-case analysis? We could and the results would probably only be slightly biased. A complete-case analysis of our radiation cohort makes the strong assumption that the missing disease status at follow-up is MCAR. In comparison, the analysis with weighted adjustment makes a weaker assumption of MAR, i.e., that the missing disease status at the follow-up is MCAR given the values of the covariates used to model the propensity of response in the follow-up.

The lack of effect of weighting in our example could be due to the fact that the weights are not highly related to the probabilities of disease outcome and therefore would not have an appreciable effect in altering the point estimates. The lack of effect on the width of the confidence intervals is partly due to the large cohort size and partly due to our effort to control for large variation of the weights by trimming. In our data the range of weights was not very wide and so the trimming did not have much of an effect. In general, one needs to be cautious in trimming the weights because trimming extreme weights to reduce variability might introduce bias and reduce the advantages of the weighting (23, 24).

Should we conduct an analysis adjusted for nonresponse and should others conducting similar studies also perform analyses weighted for nonresponse? We most certainly recommend this approach because it is a (relatively) simple method for evaluating the effect of non-response on the estimates of interest. If there is little appreciable effect, a complete-case analysis can be reported along with the information that an analysis for nonresponse did not indicate bias in the estimates. This type of information can only strengthen the study's conclusions. On the other hand, if weighted estimates are different from the complete-case method, the investigator would choose to present the presumably less biased estimates adjusted for non-response, again adding strength to a study's conclusions.

AEP Vol. 15, No. 2
February 2005: 129–136

TABLE 2. Selected standardized incidence ratios (SIRs) and confidence intervals (CIs) for the complete-case and weighted analyses among 90,302 responders to Survey 1, US Radiologic Technologist Cohort, 1983–1998

Type of cancer	Observed	SIR* (95% CI)	SIR† (95% CI)
All sites excluding NMSC[1]	3453	1.12 (1.09–1.16)	1.09 (1.06–1.12)
All solid tumors	3077	1.10 (1.06–1.14)	1.07 (1.04–1.11)
Salivary gland	8	0.98 (0.42–1.94)	0.91 (0.37–1.86)
Trachea, bronchus, and lung	307	0.86 (0.76–0.96)	0.77 (0.70–0.84)
Breast (male and female)	972	1.16 (1.09–1.24)	1.16 (1.09–1.23)
Colon cancer	210	1.11 (0.96–1.27)	1.06 (0.94–1.17)
Uterine	291	1.33 (1.18–1.50)	1.34 (1.18–1.51)
Prostate	222	1.01 (0.88–1.15)	1.02 (0.89–1.16)
Melanoma	237	1.57 (1.38–1.79)	1.59 (1.38–1.80)
Thyroid	124	1.62 (1.34–1.93)	1.61 (1.34–1.88)
Urinary bladder	84	0.92 (0.73–1.14)	0.95 (0.71–1.18)
All lymphatic and hematopoietic system	274	1.23 (1.09–1.38)	1.15 (1.00–1.31)

*Estimates calculated using only the complete cases.
†Estimates calculated using weighting.
[1]NMSC = Non-melanoma skin cancer.

The fact that adjustment for nonresponse had a small effect on the risk estimates when compared with the complete-case estimates in the current study does not necessarily mean that nonresponse bias does not exist. Nonresponse bias is a function of both the magnitude of nonresponse rate and how different the nonrespondents are from respondents. Two reasons for the small effect on the estimates due to weighting in the present analysis could be that: 1) the nonrespondents were not very different from the respondents with respect to factors that affect risk, and 2) the baseline covariates available for modeling the response propensity do not strongly predict response. To assess ability of the propensity models to predict response status, we computed the area under the receiver operating characteristic (ROC) curve for each of the models. The ROC curve is a plot of the sensitivity vs. one minus specificity for all possible cutpoints that could be used to classify individuals as respondents when the predicted probability from the logistic regression propensity models is greater than the cutpoint (27). The areas under the ROC curves were found to be 0.63 for the males and 0.59 for the females. These areas are not much larger than .5, which would result from a completely random allocation of respondent status, indicating a low level of prediction (27).

It is known that individuals who are ill from a serious disease such as cancer, stroke, or HIV can be less likely to participate in a study than healthier individuals (28–31). This can lead to nonignorable nonresponse (6) in cohort studies like the radiologic technologist study where follow-up interviews at each wave are required of study participants to determine disease incidence. In this article, we do not consider methods for adjusting for nonignorable nonresponse because results of analyses using these methods can be very sensitive to the choice of the missing data model and it is almost always impossible to validate the missing data model.

Although some cohort analyses account for missing outcomes (such as adjusting the standardized mortality ratios for unavailable death certificates) (32, 33), we are unaware of other analyses of cohort data where our proposed approach has been applied to SIR analysis. When the nonresponse rate is high or there is a reason to believe the nonrespondents differ from the respondents, it is recommended that a sensitivity analysis be performed with some adjustment for nonresponse. Although there are several alternative imputation methods and several alternative methods for weighted adjustments for nonresponse (6, 9), they tend to be more complicated to implement. The weighting method described in this article can be implemented easily using existing software packages and commonly used analysis techniques. Routine logistic regression analyses are used to obtain the response probabilities and to compute the weights for adjusting observed number of cases and accumulated person-years, and standard software is used to compute SIRs. Other software packages (SAS, SUDAAN, etc.) can also incorporate weights to obtain estimates of interest like odds ratios and rate ratios. The adjustments made for small proportion of cases are very simple to use and these formulas are available in the Appendix.

We have illustrated our method for two time points but this can be adapted to surveys with multiple time points. Different propensity weighting models can be used to compute weights for intermittent or monotone pattern missingness. A few recommendations are: 1) To use baseline information to predict response to the first follow-up and then predict response to the second follow-up conditional on response to the first follow-up using data collected at the first follow-up and so on; 2) The baseline information can be used to predict response at any point; 3) A combination of data collected at all the previous follow-ups including the baseline can be used to predict response to the follow-up of interest; and 4) When deaths are known essentially with certainty and provide complete information about the disease outcome, then the propensity of response for those individuals should be set to one. Whatever the method used to adjust for nonresponse, caution should be exercised to not greatly increase variance in the estimated measures of effect.

In summary, this article describes a novel application of an existing survey research method to adjust for nonresponse in an SIR analysis of a cohort that was queried at multiple time points. Rather than ignoring the nonrespondents, particularly if their characteristics differ from respondents, we advocate incorporating all available information to

adjust for nonresponse with the aim to improve study validity.

We are grateful to: Nathan Appel, Laura Bowen, and Henry Chen of Information Management Services for biomedical computing; Dr. Bruce Alexander of the University of Minnesota for providing initial data on respondent characteristics; and Dr. Jay Lubin for thoughtful review and helpful comments.

APPENDIX

Modifications to Account for Weighting as Well as the Small Proportion of Events

Here we show the details of the adjustment made to the Clopper-Pearson (1-α) level confidence interval, which appropriately accounts for the additional variability due to the weighting of the SIR estimator, when the observed number of cancers was greater than zero and less than 51.

Let O_w be the weighted observed number of incidence cancers and E_w be the weighted expected number of incidence cancers from a cancer registry. We are interested in computing confidence intervals for the SIR,$(\frac{O_w}{E_w})$, when the observed number of events is between zero and 51. We assume E_w to be fixed because it is based on large numbers from the registry, and calculate confidence intervals for $\hat{p}_w = (O_w/\sum_i w_i)$:

Compute the Jackknife variance of $\hat{p}_w$:

$$var_j(\hat{p}_w) = \frac{var_j(O_w)}{\left(\sum_i w_i\right)^2}$$

Proposed Method:

1. Compute effective sample size

$$N^* = \frac{\hat{p}_w(1-\hat{p}_w)}{\mathrm{var}_j(\hat{p}_w)}$$

$$\text{If, } \frac{\mathrm{var}_j(\hat{p}_w)}{\hat{p}_w(1-\hat{p}_w)/N} < 1 \text{ then } N^* = N, \text{ the sample size.}$$

2. Compute confidence intervals $(P_L(x'), P_U(x'))$

$$x' = N^*\hat{p}_w$$

$$P_L(x') = \frac{\nu_1 F_{\nu_1,\nu_2}(\alpha/2)}{\nu_2 + \nu_1 F_{\nu_1,\nu_2}(\alpha/2)}$$

$$P_U(x') = \frac{\nu_3 F_{\nu_3,\nu_4}(1-\alpha/2)}{\nu_4 + \nu_3 F_{\nu_3,\nu_4}(1-\alpha/2)}$$

where,

$$\nu_1 = 2x'$$

$$\nu_2 = 2(N^* - x' + 1)$$

$$\nu_3 = 2(x' + 1)$$

$$\nu_4 = 2(N^* - x')$$

The (1-α)-percent confidence intervals for $E(\hat{p}_w)$are given by:

$$\left(\left(\sum_i w_i\right)P_L\left(x'\right), \left(\sum_i w_i\right)P_U\left(x'\right)\right)$$

REFERENCES

1. Steeh CG. Trends in nonresponse rates, 1952–1979. Public Opin Q. 1981;45:40–57.
2. Kessler RC, Little RJA, Groves RM. Advances in strategies for minimizing and adjusting for survey nonresponse. Epidemiol Rev. 1995;17(1):192–204.
3. Marmor JK, Oliveria SA, Donahue RP, Garrahie EJ, White MJ, Moore LL, et al. Factors encouraging cohort maintenance in a longitudinal study. J Clin Epidemiol. 1991;44(6):531–535.
4. Blumenthal DS, Sung J, Coates R, Williams J, Liff J. Recruitment and retention of subjects for a longitudinal cancer prevention study in an inner-city black community. Health Serv Res. 1995;30:197–205.
5. Hunt JR, White E. Retaining and tracking cohort study members. Epidemiol Rev. 1998;20(1):57–70.
6. Little RJA, Rubin DB. Statistical Analysis with Missing Data. New York, NY: John Wiley & Sons, Inc.; 2002.
7. Robbins JM, Rotnitzky A, Zhao LP. Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. Journal of the American Statistical Association. 1995;90:106–121.
8. Preisser JS, Galecki AT, Lohman KK, Wagenknecht LE. Analysis of smoking trends with incomplete longitudinal binary responses. Journal of the American Statistical Association. 2000;95:1021–1031.
9. Schafer JL. Analysis of Incomplete Multivariate Data. UK: Chapman & Hall; 1997.
10. Kalton G. Handling wave nonresponse in panel surveys. J Off Stat. 1986;2:303–314.
11. Lepkowski JM. Treatment of wave nonresponse in panel surveys. In: Kasprzyk D, Duncan G, Kalton G, et al, eds. Panel Surveys. New York, NY: John Wiley & Sons, Inc.; 1989:348–374.
12. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika. 1983;70:41–55.
13. Little RJA. Survey nonresponse adjustments for estimates of means. Int Stat Rev. 1986;54:139–157.
14. Goksel H, Judkins DR, Mosher WD. Nonresponse adjustment for a telephone follow-up to a national in-person survey. Proceedings of the Survey Research Methods Section, American Statistical Association. 1991:581–586.
15. Woodburn L. Using auxiliary information to investigate nonresponse bias. Proceedings of the Survey Research Methods Section, American Statistical Association. 1991:278–283.
16. Czajka JL, Hirabayashi SM, Little RJA, Rubin DB. Projecting from advance data using propensity modeling: An application to income and tax statistics. J Bus Econ Stat. 1992;10:117–131.
17. Korn EL, Graubard BI. Analysis of Health Surveys. New York, NY: John Wiley & Sons, Inc.; 1999; 29–31, 164, 172–176.
18. Sigurdson AJ, Doody MM, Rao RS, Freedman DM, Alexander BH, Hauptmann M, et al. Cancer Incidence in the US Radiologic Technologists Health Study, 1983–1998. Cancer. 2003;97:3080–3089.
19. Doody MM, Mandel JS, Lubin JH, Boice JD Jr. Mortality among United States radiologic technologists, 1926–1990. Cancer Causes Control. 1998;9:67–75.
20. Mohan AK, Hauptmann M, Linet MS, Ron E, Lubin JH, Freedman DM, et al. Breast cancer mortality among female radiologic technologists in the United States. J Natl Cancer Inst. 2002;94(12):943–948.

21. Boice JD Jr, Mandel JS, Doody MM, Yoder RC, McGowan R. A health survey of radiologic technologists. Cancer. 1992;69:586–598.

22. Doody MM, Sigurdson AS, Kampa D, Chimes K, Alexander BH, Ron E, et al. Randomized trial of financial incentives for improving response to a mailed questionnaire. Am J Epidemiol. 2003;157:643–651.

23. Potter FJ. A study of procedures to identify and trim extreme sampling weights. Proceedings of the Survey Research Methods Section, American Statistical Association. 1990:225–230.

24. Elliott MR, Little RJA. Model-based alternatives to trimming survey weights. J Off Stat. 2000;16:191–209.

25. Clopper CJ, Pearson ES. The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrika. 1934;26:404–413.

26. Liddell FDK. Simple exact analysis of the standardized mortality ratio. J Epidemiol Community Health. 1984;38:85–88.

27. Hosmer DW, Lemeshow S. Applied Logistic Regression. New York, NY: John Wiley & Sons, Inc.; 2000:160–164.

28. Rockwood K, Stolee P, Robertson D, et al. Response bias in a health status survey of elderly people. Age Ageing. 1989;18(3):177–182.

29. Heilbrun LK, Ross PD, Wasnich RD, Yano K, Vogel JM. Characteristics of respondents and nonrespondents in a prospective study of osteoporosis. J Clin Epidemiol. 1991;44(3):233–239.

30. Mugusi F, Josiah R, Moshi A, Chale S, Bakari M, Aris E, et al. Dropouts in a long-term follow-up study involving voluntary counseling and HIV testing: Experience from a cohort of police officers in Dar Es Salaam, Tanzania. J Acquir Immune Defic Syndr. 2002;30(1):119–123.

31. Duncan PW, Reker DM, Horner RD, Samsa GP, Hoenig H, LaClair BJ, et al. Performance of a mail-administered version of a stroke-specific outcome measure, the Stroke Impact Scale. Clin Rehabil. 2002;16:493–505.

32. Rittgen W, Becker N. SMR analysis of historical follow-up studies with missing death certificates. Biometrics. 2000;56(4):1164–1169.

33. Blettner M, Zeeb H, Langner I, Hammer GP, Schafft T. Mortality from cancer and other causes among airline cabin attendants in Germany, 1960–1997. Am J Epidemiol. 2002;156(6):556–565.